IMAGE TRANSMISSION FOR LOW BANDWIDTH WITH REGION OF INTEREST

FIELD OF THE INVENTION

[0001] The present invention relates to transmitting image data for display on two different display screen areas, and more particularly, this disclosure provides a system and method for transmitting image data for display between two differently sized display areas such that a region of interest is viewable to a user.

BACKGROUND OF THE INVENTION

[0002] Handheld or portable devices, such as personal digital assistants (PDAs), provide a user the flexibility of remotely accessing information sources such as servers, databases and the internet. They are also used to operate software that was originally designed for desktop machines. For example, many Microsoft desktop applications will run on handheld machines. One of the main drawbacks that is encountered when using a handheld device to access these types of information sources is that the image data is not adapted to the smaller display area of the handheld device. In other words, the image does not "fit" within the handheld device's smaller display area.

[0003] According to one technique developed to overcome the problems caused by the disparity between the handheld device display area size to the actual image size, only a portion of the image is displayed within the smaller display area. The user can then navigate through the entire image using the hand held devices input interface (e.g., wand, keys, buttons, mouse, etc.) to view the other portions of the image. For instance, if a user remotely accesses a web page with a PDA, only a small portion of the web page is displayed within the PDA display area and the user navigates to other portions of the web page via the user interface. This partial viewing of the image can be confusing and frustrating to the user since it can become unclear where in the image the user is viewing, thereby making navigation through the image difficult. Hence, partial viewing represents one of the major impediments to interacting with information intended for large display devices. One alternative to partial viewing is

that the full sized image is "scaled down" into the smaller display area of the small display area device. However, this technique often makes it difficult or impossible to view image details depending on the size of the display area.

[0004] In another known technique, instead of equally "scaling down" the whole image, the entire image is displayed in a PDA-type display area by scaling down a selected portion (or region of interest) at a relatively low scaling factor while using a relatively higher scaling factor or factors for the remainder of the image. In this way, the region of interest is readable by the user and the remainder of the image is discernable enough to provide context for navigation. When the user navigates to other portions of the full image with a user interface and selects new regions of interest, the scaling factor of the new region of interest is reduced (i.e., reduced such that the region becomes readable) while the remainder of the image regions including the old region of interest is scaled down using a higher scaling factor. This technique facilitates navigation about a larger image that is displayed on a relatively smaller display area.

[0005] In order to display a scaled version of an image as described above, one known method (shown in Fig. 1A) is to convert the image region into a frequency domain representation including a plurality of frequency coefficients and then transmit all the frequency coefficients from the image source to the destination device (e.g. a PDA). Next, the transformation is inverted to obtain a full sized spatial domain image region, and then this image region is scaled down to the desired size in the spatial domain. This method is illustrated in Fig. 1A, such that an image region is converted into a frequency domain representation and a frequency coefficient array 10 is transmitted to the destination device. The inverse transformation converts the array back to spatial image data 11 of the same size as the original image region. Then image region 11 is scaled to the desired size 12. The problem with this method is that it requires the transmission of all frequency coefficients and hence, requires a large amount of bandwidth.

[0006] The disadvantage of the method shown in Fig. 1A is that whenever remotely accessing image sources with a handheld device the image data is often wirelessly

2

transmitted over a communication path having a finite bandwidth. A small or overburdened communication path can significantly hinder the users ability to quickly access remote information by limiting the amount of data that can be transmitted between the information source and the handheld device. Hence, there is a strong motivation to reduce bandwidth usage while providing as much information as possible. However, transmitting all frequency coefficients tends to consume a relatively large amount of bandwidth that can potentially result in slow remote access times or an inability to remotely access certain image data. Hence, what is needed is a manner in which to transmit digital images between an image source and a handheld device in a bandwidth efficient manner.

[0007] The present invention, in general, is directed towards displaying large images and user interfaces originating from an information source (e.g., a server) within a small display area, such as the display area of a handheld device (e.g., PDA). In addition, this disclosure is directed to providing a system and method of transmitting an image in a bandwidth conservative manner from an information source to a handheld device.

SUMMARY OF THE INVENTION

[0008] A system and method of transmission is described for transmitting image data adapted to a first display area size to an apparatus having a smaller second display area size. The system and method of transmission allows a user to view a large image on a hand held device with a small display area by magnifying a region of interest within the large image and de-magnifying peripheral areas about the region of interest, causing, for example, a "fisheye" view of the image to be displayed in the handheld device display area.

[0009] In accordance with the system and method, the image data, which is adapted for display in a first display area size, is divided into a plurality of regions each having a corresponding block of image data. Each block of image data is then converted into a frequency domain representation of the block having an associated array of coefficients. Frequency coefficients are cropped from at least a portion of the arrays so as to spatially scale corresponding image regions within the spatial domain dependent

3

on the amount of cropping. Then cropped and uncropped arrays are transmitted to the apparatus for displaying a scaled version of the image in the second display area. In one embodiment, coefficient arrays of the image regions corresponding to a region of interest are uncropped and coefficient arrays of image regions peripheral to the region of interest are cropped. The coefficients corresponding to the region of interest, together with the cropped portions of coefficients for the peripheral regions, are transmitted to the apparatus. These transmitted coefficients are reverted to the spatial domain to render the "fisheye" view in the second display area.

Docket No.: 10018132-1

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] The objects, features, and advantages of the present invention will be apparent to one skilled in the art, in view of the following detailed description in which:

[0011] **Fig. 1A** a prior art method of scaling image data transmitted from an image source having a first associated display area size to a destination device having a second smaller display area size;

[0012] **Fig. 1B** shows examples of different remote images being wirelessly transmitted to a handheld device;

[0013] **Fig. 2** illustrates a method of transmitting image data adapted to a first display area to an apparatus having a second display area according to the present invention;

[0014] **Fig. 3** shows an example of an image being divided into a plurality of regions;

[0015] **Fig. 4** illustrates the concept of cropping frequency coefficients corresponding to an image region to obtain a spatially scaled image region;

[0016] **Fig. 5** illustrates coefficient cropping in the frequency domain representation of a full image and its effect on the full image in the spatial domain;

[0017] **Figs. 6A and 6B** show an example of scaling a full size image by scaling at least a portion of the blocks of the image data using coefficient cropping;

[0018] **Fig. 6C** shows a scale map for the image shown in Fig. 6A;

[0019] **Fig. 7A** illustrates an embodiment of a system for transmitting image data adapted to a first display area to an apparatus having a second smaller display area according to the present invention;

[0020] **Fig. 7B** illustrates an embodiment of an apparatus for receiving and displaying image data transmitted from the system shown in Fig. 7A;

[0021] **Fig. 8** illustrates a second embodiment of the method for transmitting image data adapted to a first display area to an apparatus having a second smaller display area having a changing region of interest; and

[0022] **Fig. 9** illustrates a second embodiment of a system for transmitting image data adapted to a first display area to an apparatus having a second smaller display area;

[0023] **Figs. 10A and 10B** illustrate an embodiment of the method for transmitting image data in a prioritized and progressive manner, based on perceptual significance.

[0024] **Fig. 11** illustrates an embodiment of the method for color space sub-sampling to improve bandwidth efficiency.

DETAILED DESCRIPTION

[0025] In general, the system and method of the present invention is a technique of transmitting an image adapted to a first display area size, such as a standard computer screen having a particular pixel-by-pixel resolution, to an apparatus having a smaller display area size by scaling the image in the spatial domain by means of coefficient cropping in the frequency domain. When the cropped coefficients are used to display an image within the second smaller display area, a scaled version of the image is displayed. In a specific embodiment, frequency domain coefficients are cropped such that a region of interest of the image is scaled down less than the remainder of the image when displayed in the second smaller display area. Scaling in this manner provides the user with a readable region of interest and with the remainder of the image scaled down so as to facilitate easy image navigation by the user. Furthermore the system and method is a technique of transmitting image data to the apparatus in a bandwidth efficient manner such that only a required subset of the coefficient data is transmitted and combined with previously transmitted coefficient data when the region of interest is changed.

[0026] Fig. 1B shows examples of different images adapted to a first display area size that might be transmitted to an apparatus 13 having a display area 14 of a second smaller display area size. As shown in the examples, the image can include a webpage image, the image of the desktop of a users personal computer (PC) that is being remotely accessed, or the image of a portion of a database. Any of these images can be transmitted over an intranet, internet, or direct connection. In addition, the images may be wirelessly transmitted or transmitted over a wire connection (not shown). The images may be accessed through a server or may be directly accessed. The apparatus 13 may be any device having at least a display area, a means of receiving an image signal, and a means for driving the display area to display an image corresponding to the image signal within the display area. Some commonly known examples of these types of apparatus include a PDA, a telephone, a cellular phone, and a laptop.

7

[0027] Fig. 2 shows a first embodiment of a method of transmitting an image adapted to a first display area size to an apparatus having a second display area size which includes dividing the image into a plurality of regions each having a corresponding block of data (block 20). For instance, Fig. 3 shows an example of an image 30 being divided into a plurality of regions, A-I, each region having a corresponding block of data of $n \times n$ pixels.

[0028] Each of the blocks of data is converted into a frequency domain representation of the data to generate a corresponding array of frequency coefficients (block 21, Fig. 2). Hence, each block of data corresponding to each of regions A-I in the example shown in Fig. 3 is converted into a corresponding array of frequency coefficients.

[0029] According to the present invention, the task of spatial scaling is achieved by frequency coefficient cropping. Hence, unlike the prior art method shown in Fig. 1A in which the scaling of each region is performed within the spatial domain after transmitting to a destination device, the method shown in Fig. 2 crops the frequency array of the region to be scaled such that once it is reconverted into the spatial domain it is already scaled to the desired factor. The effect of frequency cropping a given image in its frequency domain representation followed by a transformation to the spatial domain is mathematically equivalent to and yields essentially the same results as the prior art technique of spatially scaling the same given image. Hence, according to the method shown in Fig. 2, frequency coefficients of a portion of the arrays are cropped so as to scale at least a portion of corresponding regions of the image within the spatial domain (block 22) and then blocks of cropped coefficients and any remaining uncropped coefficient arrays are transmitted to the apparatus for displaying a scaled version of the image in the second display area (block 23).

[0030] Fig. 4 illustrates the concept of cropping frequency coefficients to obtain a spatially scaled image region. In the more bandwidth efficient manner according to the present invention, a frequency coefficient array 40 corresponding to an image region is first cropped to the desired scale 40A determined by the display size on the client device. Only the necessary coefficients 40A are transmitted to the client device and reconverted to produce an image region 41 scaled proportionally to the amount of

8

coefficients that have been cropped. Note that image region 41 looks the same as the prior art scaled image region 12 (Fig. 1A). When an image of a larger scale is desired, only supplementary frequency coefficients 40B need to be transmitted. The supplementary coefficients 40B can be combined with previously transmitted coefficients 40A to generate an image region of the original size42. Note that we can further enlarge the image by padding zeros to the transmitted frequency coefficients to larger scales.

[0031] The method shown in Fig. 4 as applied to a full image is shown in Fig. 5. First, the full image is divided into regions 51. Secondly, each region is converted to an array of frequency coefficients 52. These frequency coefficients are then cropped to the desired scales 53. Only necessary coefficients 54 are transmitted to the client apparatus. An inverse transformation yields the second display of desired scales 55. Note that, in the embodiment shown in Fig. 5, the frequency coefficients associated with the region of interest (ROI) are not cropped, thus after transmission and inverse transformation it is restored to the original size. As mentioned above, the ROI can be further enlarged by padding zeros to the transmitted frequency coefficients associated with the ROI.

[0032] Hence, only necessary coefficients are transmitted to the apparatus, where an inverse transformation generates the second display with regions scaled proportionally to the extent of cropped frequency coefficients. More importantly, if afterwards a larger image is desired, only supplementary frequency coefficients need to be transmitted. These supplementary coefficients can be combined with previously transmitted coefficients to render a less scaled image. This scheme is bandwidth efficient as only necessary data is transmitted, and all transmitted data can be reused.

[0033] Figs. 6A-6C show an example of a full size image scaled to be displayed on a second smaller display area. Fig. 6A shows a full size image partitioned into regions each having associated blocks of image data. Fig. 6B shows a scaled version of the image. As can be seen some regions of the scaled image have higher scaling than others. For instance, the corner regions are scaled greater than the side regions and the side regions are scaled greater than the center region. In this example, the center

9

region (i.e., the region of interest) has been scaled, so as to make it readable to the user, however, the center region need not be scaled at all and may be displayed at full size. Fig. 6C shows a possible scale factor mapping of the image where the data blocks corresponding to the corners have been scaled by a factor of 8:1, the data blocks corresponding to the sides have been scaled by a factor of 4:1, and the data block corresponding to the center has been scaled by 2:1.

[0034] Fig. 7A illustrates one embodiment of a system 70 for transmitting image data adapted to a first display area size to a second smaller display area size including an image divider 71 for receiving image data 70A that is adapted to a first display area size and for dividing the image into a plurality of regions each corresponding to a block of image data 71A. The system further includes an image data converter 72 for converting each of the plurality of blocks of data 71A into a corresponding frequency domain representation 72A – each representation including an array of coefficients. A frequency coefficient cropper 73 functions to crop arrays of coefficients depending on the desired scale to generate blocks of cropped coefficients 73A. A transmitter 74 transmits the cropped blocks of coefficients 73A as well as any uncropped arrays 72A. In one embodiment, the transmitter includes an encoder 74B for compressing the coefficient data 73A and 72A prior to transmission. In another embodiment, frequency coefficient cropping is performed dependent on a selected region of interest (i.e., ROI) pointer signal.

[0035] Fig. 7B illustrates an apparatus 75 for displaying a scaled version of the image data 70A including a receiver 76 for receiving the cropped frequency coefficients 74A. In one embodiment, the receiver 76 includes a decoder 76B for decoding the compressed frequency data 74A. The receiver 76 provides decoded image data 76A to an image data inverter 77 to transform the frequency coefficients back to image data 77A. The image data 77A is passed to a display driver 78 which processes the image data to be displayed in the second display area 79. In one embodiment, the apparatus includes a user interface 79B for identifying a region of interest within the second display area and a means of transmitting a ROI signal to system 70.

10

[0036] In another embodiment of the system and method of the present invention, a new region of interest can be identified within a currently displayed scaled version of the image and supplemental frequency coefficient data can then be transmitted to the apparatus and combined with previously transmitted frequency coefficient data so as to display the scaled version of the image dependent on the new region of interest without having to re-transmit the coefficient data for the full image. The supplemental frequency coefficient data represents the difference between previously transmitted coefficient data including cropped and un-cropped blocks of coefficients, and newly determined coefficient data including new cropped and un-cropped blocks of coefficients determined in view of the new region of interest.

[0037] Fig. 8 shows one embodiment of the method of transmission in which a new region of interest is identified. According to this embodiment, an image is divided into regions including a region of interest. Each region having a corresponding block of data (block 80), each block of data is converted into a frequency domain representation (block 81), and at least a portion of the frequency domain representations of the blocks are cropped dependent on the region of interest (block 82) so as to scale the image in the spatial domain. The coefficient data including the cropped blocks and uncropped blocks is then transmitted to the apparatus for displaying a scaled version of the image within the second display area (blocks 83). According to this method, frequency coefficients are cropped dependent on the region of interest such when the scaled version of the image is displayed in the second display area the region of interest is "readable" by the user and the non-regions of interest are discernable by the user so as to enable navigation through the image.

[0038] Once the coefficients are transmitted, a new region of interest can be identified (block 84). Identification of the region of interest can be performed by the user selecting the region within the second display area with a user interface adapted for interfacing with the apparatus. For example the user may use a "wand" to point to the second display area screen that is responsive to tactile user input. Alternatively, the user may select a new region of interest using buttons on the apparatus for navigating within the scaled image displayed within the second display area.

[0039] New coefficient data is determined including new cropped blocks of coefficients and new un-cropped blocks of coefficients dependent on the identified new region of interest (block 85). These cropped blocks and uncropped blocks of coefficients are determined in the same manner as the previously determined cropped and uncropped coefficients. Specifically, the frequency domain representation of the plurality of blocks of data corresponding to regions of the image are cropped dependent on the new region of interest such that the region of interest of the image is scaled so as to be readable by the user and the non-regions of interest are scaled so as to be discernable enough to enable navigation through the image.

[0040] In order to reduce bandwidth needed to transmit the scaled image to the apparatus according to this embodiment of the method of transmitting, a subset of coefficients are determined (block 86) that represents the difference between previously transmitted coefficient data (block 83) and the new coefficient data.

[0041] The subset of coefficients is transmitted (block 87) to the apparatus to be combined with the coefficient data that had been previously transmitted to the apparatus in order to display the scaled version dependent on the new region of interest. In other words, the subset of coefficient data transmitted to the apparatus corresponding to the new region of interest is used to essentially update and compliment the buffer storing all previously transmitted coefficient data. As new regions of interest are selected/changed, additional supplemental coefficients are transmitted.

[0042] In accordance with the method shown in Fig. 8, the system 70 as shown in Fig. 9 further includes a first buffer 90 for storing the previously transmitted coefficient data and a second buffer 91 within the apparatus 75 for storing all previously transmitted coefficient data received by the apparatus. The size of buffer 91 is such that it is able to store all coefficient data for an image formatted for displaying within the first display area (i.e., a full resolution image). The second buffer accumulates the coefficients until it potentially has a complete uncropped set of blocks of coefficients corresponding to the image. The first and second buffers are synchronized such that there is data consistency between the buffers (i.e., the first buffer as stored by the

12

system is a copy of the second buffer as stored by the receiver). In one embodiment, the apparatus 75 includes a buffer synchronization mechanism (not shown) that interacts with the server to ensure that the buffers are synchronized.

[0043] In another embodiment, a bitstream of the encoded image signal 74A is transmitted by system 70 such that coefficient data 90A encoded within signal 74A is received by the apparatus 75 in an order of most perceptually significant to least perceptually significant. The perceptually significant sequence of transmitting the encoded coefficient data begins with transmitting the low order (i.e., low frequency) frequency coefficients so that a baseline image can be displayed and can be further refined. Next, higher order (i.e., high frequency) coefficients of the ROI are transmitted to further improve the ROI display quality. Then, the periphery regions about the ROI are further refined by transmitting the remainder of the coefficient data. When all visible information has been sent for the current view, the supplemental coefficients of the periphery can be sent if additional bandwidth is available. This is a form of predictive caching. If the user changes the ROI, the high order coefficients will become useful. If the user changes the ROI in a predictable fashion, by following down a column of text for example, the next ROI can be extrapolated and additional coefficients can be sent to refine the future ROI. This procedure is useful for applications requiring prompt feedback and responsiveness. It should be noted that the bitstream may be terminated at any time, for example, when the region of interest changes thereby initiating the generation of a new bitstream dependent on the new region of interest.

[0044] The system 70 further includes a ROI pointer register 93 for storing the region on interest point sent from the receiver as identified by the apparatus' user interface which is used to determine how each region of the image is scaled and hence the cropping of each corresponding coefficient array.

[0045] In one embodiment, the receiver 75 further includes an image renderer 94. In response to the ROI pointer provided by the user interface 79B, a rendering algorithm identifies the coefficients in the second buffer 91 to be used for displaying the scaled version of the image within the second display area.

13

[0046] In yet another embodiment, the perceptually significant sequence of transmitting the encoded coefficient data 74A is compiled in a prioritized and progressive manner. Once the user selects a region of interest, for each frequency coefficient block, the number of coefficients necessary for rendering the scaled second display 90 is determined, and the number of coefficients that are already transmitted is indicated in an image mask stored within system 70. Two statistics are constantly evaluated to prioritize the coefficients, one is the percentage of necessary coefficients that are already transmitted, the other is the number of necessary coefficients that are yet to be sent. The highest priority goes to the frequency coefficient block with the lowest percentage of necessary coefficients that are already transmitted. If all frequency coefficient blocks have the same percentile, then the highest priority goes to the frequency coefficient block with the most number of necessary coefficients that are yet to be transmitted. If all frequency coefficient blocks are still equal, then the coefficients are chosen following raster order, from left to right, from top to bottom. Within each frequency coefficient block, the coefficients are ordered in a zigzag manner similar to that is commonly used in JPEG convention.

[0047] Figs. 10A and 10B illustrate the above prioritized coefficient transport scheme. Frequency coefficient blocks 101A-103A (Fig. 10A) represent three blocks of coefficients for transmission having different starting conditions such as necessary coefficients, percent transmitted, percent to be transmitted, etc. The filled areas in frequency coefficient blocks 101A-103A indicate non-zero coefficients; the shaded areas of the Scaled blocks 101B-103B denote the coefficients necessary for rendering the scaled view in the second display area; whereas the shaded areas in the Mask blocks 101C-103C represent coefficients that are already transmitted or coefficients that are zero. Note that all coefficients are zero by default, therefore all zero coefficients are considered to have been transmitted by default and appear as shaded in the Mask blocks. To prioritize transmission of coefficients in the three frequency coefficient blocks 101A-103A, the statistics of each block are evaluated constantly during the ongoing process of determining which coefficients are to be transmitted, as shown in Table 1:

Docket No.: 10018132-1

	Necessary Coefficients (#)	Round	Transmitted Coefficients (#)	Transmitted Coefficients (%)	Un-transmitted Coefficients (#)
Block 101	12	0	4	33.3	8
		1	5	41.7	7
		2	5	41.7	7
		3	6	50	6
Block 102	12	0	5	41.7	7
		1	5	41.7	7
		2	5	41.7	7
		3	5	41.7	7
Block 103	9	0	3	33.3	6
		1	3	33.3	6
		2	4	44.4	5
		3	4	44.4	5

Table 1

(where a "round" of transmission corresponds to an ongoing evaluation process of the statistics for determining which coefficients are to being transmitted). Fig. 10B shows the state of the Mask Image data 101C-103C after three rounds of prioritized transmission (rounds 0-2) of coefficients of corresponding frequency coefficient blocks 101A-103A. Initially at round 0, block 101 and block 103 both have the lowest percentage of previously transmitted coefficients (33.3%), but block 101 has more un-transmitted coefficients (i.e., 8 un-transmitted coefficients), so block 101 is given the highest priority and coefficient 104 is transmitted. At round 1, after a coefficient from block 101 is selected for transmission, block 103 now has the lowest percentage of transmitted coefficients (33.3%), thus the highest priority and coefficient 105 is transmitted. Now that a coefficient from block 103 is selected to the bit stream for transmission, at round 2, block 101 and block 102 have the lowest percentage of transmitted coefficients (41.7%). To break the tie, the number of un-

transmitted coefficients is evaluated. Coincidentally, both block 101 and block 102 have the same number of un-transmitted coefficients (i.e., 7 un-transmitted). At this stage, raster order takes precedence. Since block 101 is on above block 102, block 101 gets higher priority and coefficient 106 is transmitted. At round 3, block 102 has the highest priority due to the lowest percentage of transmitted coefficients (41.7%). This process goes on until all necessary coefficients are transmitted. If more bandwidth is available, predictive caching is carried out. This transport scheme allows graceful degradation in that if the transmission channel is temporarily disabled or bandwidth availability is significantly reduced when displaying a new region of interest, the rendering engine for displaying the scaled image on the second display can still generate a low resolution version of the image using the currently available (i.e., previously transmitted) frequency coefficients. Then, as the communication channel is restored and/or bandwidth increases and more coefficients are transmitted, the image is rendered at a higher resolution.

[0048] In still another embodiment of the present invention, bandwidth efficiency can be further improved by exploiting the type of color space used to represent the image. As is well known in the field of image processing, an image block 11 (Fig. 11) can be represented in a YC_rC_b color space, which comprises a luminance channel (block 112) and two chrominance components (blocks 113 and 114). Note that high frequency information is only retained in the luminance channel 112, but not in the chrominance channels 113 or 114. This characteristic allows a more compact representation of the image prior to transmission where the chrominance channels (i.e., blocks 113 and 114) are sub-sampled by a factor of 2 to obtain blocks 115 and 116 (Fig. 11). Using luminance information 112 and sub-sampled chrominance information 115 and 116, achieves a compression ratio of 2 compared to using full color channels.

[0049] In still another embodiment of the present invention, frequency coefficients are prioritized for transmission by setting the luminance information at a higher priority for transmission than chrominance information. In this case, coefficients from the luminance channel are transmitted first to yield a monochrome display, and then coefficients from chrominance channels are transmitted to render a color display.

16

[0050] In still another embodiment, the coefficient data 90A is encoded by assuming that the system 70 and the apparatus 75 are synchronized under a guaranteed delivery protocol such that at any point of time both the system and the apparatus know which coefficients in each region of the image have been transmitted, and which additional coefficients are required. For example, if initially, coefficients for a given region in a 3x4 low-order window are transmitted to the apparatus and then the apparatus indicates (via a new ROI) that a larger window having coefficients in a 5x7 low order window is required, the system will transmit an encoded bitstream corresponding to the additional coefficients required to enable the apparatus to obtain all coefficients in 5x7 low-order window. In this embodiment, both the system and the apparatus maintain a binary mask of coefficients already known for each image region. The apparatus requests low-order $M \times N$ coefficients of which K are already known, leaving $MN - K$ unknown coefficients that must be conveyed. At the system, the quantized frequency coefficients are scanned in a zigzag order similar to that in JPEG, but excluding the ones already known or outside the requested $M \times N$ window. In this ordered set (with $MN - K$ coefficients), the position of the last non-zero coefficient, say L, is first conveyed using $\lceil \log_2 (MN - K) \rceil$ bits. L is in effect the total number of coefficients to scan out of a maximum of $MN - K$. The remaining $MN - K - L$ coefficients are dropped from the scan and reconstructed as 0.

[0051] Next, the sequence of L unknown coefficients in the zigzag scan order is conveyed using combined arithmetic and binary encoding. Each coefficient is assigned a category, based on its true value. Category 0 stand for value {0}; category 1 stands for value {-1, 1}; category 2 stands for values {-3, -4, 3, 4}; category 3 stand for values {-5, -6, -7, -8, 5, 6, 7, 8}; and so on. In general, for category c, an additional c bits specify the exact value of the coefficient. In the preferred encoding scheme, the category is arithmetic encoded, while the additional bits for each coefficient are left uncoded.

[0052] The category encoding follows several probability models, depending on the position of the coefficient being encoded, and its context. First, each frequency coefficient in zigzag scan order is classified based on its position into 7 different classes as depicted in Table 2 below.

Docket No.: 10018132-1

0	1	2	3	3	4	4	5
1	2	3	3	4	4	5	5
2	3	3	4	4	5	5	6
3	3	4	4	5	5	6	6
3	4	4	5	5	6	6	6
4	4	5	5	6	6	6	6
4	5	5	6	6	6	6	6
5	5	6	6	6	6	6	6

Table 2

Then, based on its context (depending on neighboring coefficients already encoded), the class is moved up or down. For example, consider a coefficient at the fourth column of the third row. Then based on position alone, the class of the coefficient is 4 (see Table 2). However, if from its context it is found that the current coefficient is likely to have a larger (smaller) value, the class is modified to a lower (higher) class, such as 3 (5), if there is room. In particular, for each coefficient, the context is given by the categories of one or more neighboring coefficients that come before the current coefficient in scan order.

[0053] The modified class of the coefficient yields one of 7 probability models to be used for arithmetic encoding of the coefficient's category. The models are chosen such that in higher classes, the lower categories are more probable than higher ones, but less so if the context is high. Note that 'category' depends on the true value of the coefficient, whereas 'class' determines its expected magnitude based on its position and context.

[0054] In the preceding description, numerous specific details are set forth, such as encoding techniques and the number of regions of the image in order to provide a thorough understanding of the present invention. It will be apparent, however, to one skilled in the art that these specific details need not be employed to practice the present invention. In other instances, well-known transmission steps and image processing techniques have not been described in detail in order to avoid unnecessarily obscuring the present invention.

[0055] In addition, although elements of the present invention have been described in conjunction with certain embodiments, it is appreciated that the invention can be

18

implement in a variety of other ways. Consequently, it is to be understood that the particular embodiments shown and described by way of illustration are in no way intended to be considered limiting. Reference to the details of these embodiments is not intended to limit the scope of the claims which themselves recited only those features regarded as essential to the invention.